MEDIA MATTERS: DEVELOPING A MEDIA DRIVEN MODEL TO UNDERSTAND BILL PASSAGE RATE IN THE ARIZONA STATE LEGISLATURE

by
Benjamin Roth

A capstone project submitted to Johns Hopkins University in conformity with the requirements for the degree of Master of Science in Data Analytics for Policy

Baltimore, Maryland
December 2020

<u>*Abstract*</u>

The influence of media effects on policy is well documented. Existing research often points to clear relationships between media attention on national issues and legislative activity on those issues; increased attention on a policy area leads to an increase in bills passed on that topic. While state legislatures pass bills at a higher rate than Congress, less research has been focused on state-level activity. As state-based newspapers often focus on their state-relevant issues more so than national newspapers, this research seeks to develop a generalized media effects model by focusing on article activity of four Arizona newspapers: The Arizona Daily Star, the Arizona Republic, the Arizona Capitol Times, and the Phoenix New Times. While accounting for bill meta variables (number of sponsors, number of committees introduced to, etc.) and bill text data, a logit model was trained on per-bill-topic article counts from each of those four newspapers to determine their effect on the probability of bill passage in the Arizona Legislature. None of the article variables proved to be significant, suggesting that local media coverage on the topic of general legislative topics does not have the same impact on bill passage as national coverage on singular issues. Bill passage rate, however, positively statistically significantly increased with increases to bill meta variables; this finding confirms that these demographic variables are important in understanding why a bill passes. Finally, while initial analysis showed that text data is not inferential to the passage of bill, more work remains to confirm this finding.

Table of Contents

1. Introduction

Recent years have seen steady viewership in all of local, cable, and network news, despite growing popularity of online media.[1] While media attention has continued, congressional legislative inaction has persisted. From 2013-2014, Congress only passed 352 bills and resolutions; state legislatures and Washington D.C. passed 45,564 bills and resolutions over the same time period.[2] Although state legislative activity clearly has a larger impact on citizens than that of the federal government, studies have found that Americans have minimal knowledge of their own state governments.[3] Given the salience of traditional media and the importance of state-level legislation, a clear opportunity for unification presents itself. This research seeks to understand the extent to which media attention on a bill topic impacts the probability of that bill's passage in a single state: Arizona.

Existing research has thoroughly documented the impact of media on legislative activity. Whether this effect has been observed directly, indirectly through public opinion, or complexly modeled, media coverage's influence on bill passage is evident. Any attempt to model bill passage, however, must include additional variables. In line with existing research, this study also incorporates text data and bill meta-data such as number of sponsors, percentage of sponsors in the majority party, etc. Previous studies have largely focused on singular policy areas; this research attempts to break from that traditional model by studying generalized legislation.

1 Pew Research Center, "Local TV News Fact Sheet." "Cable News Fact Sheet."; "Network News Fact Sheet."
2 Justice, "States Six Times More Productive Than Congress "
3 Rosen, "Americans Don't Know Much about State Government, Survey Finds"

The ultimate results of this research's modeling efforts reveal several key findings. State-level media attention, quantified by per-bill-topic article counts from four different Arizona news sources, was found to have non-significant effects on the passage rate of a bill. Bill meta-variables in this model, however, were shown to positively impact the rate of bill passage, agreeing with existing theory. Though there is room to increase the complexity of the analysis, text data initially shows to have no statistical impact on the passage rate of a bill.

2. Literature Review

H1: Bills introduced in the Arizona State Legislature whose topics are more covered by Arizona-based news organizations are more likely pass.

2.1 A Policy-Making Triad

The media's role in the public policy process is well researched. Serving as an intermediary between the public and the policymakers, the media helps bridge the information gap between the two groups.[4] This three-way relationship has been a focus of political science research for decades, with researchers seeking to understand if and to what extent media, public opinion, and policymaking influence each other.

2.1.1 News As a Driver of Opinion…

The news media serves as an important actor in simply determining what policy issues are worthy of importance, and consequently on which issues the public develops opinions. This topic's importance, however, is not necessarily consistent from source to source; different sources place different priorities in the amount they cover one issue or another.[5] While the news media can guide the public's attention towards a certain policy

4 Kingdon, "Agendas, Alternatives, and Public Policies", 60
5 McCombs, Maxwell E. and Donald L. Shaw. "The Agenda-Setting Function of Mass Media."

issue, the media actually may also serve as a guide to policymakers in certain instances.[6] News organizations' guidance on issues has not always been shown to see a boon. With regards to health issues, qualitative analysis has found that TV coverage of health policy may be a detriment; these consequences include the presentation of conflicting views that may lead to public confusion and distort views of certain health policy realities.[7]

While media attention on an issue brings that issue to the spotlight, the media's framing of an issue may shape the public's opinion of that issue. News coverage often varies between topics that have been in spotlight for a long time and events that merit instantaneous need for attention. Whether the news covers an immediate matter or a long discussed subject, framing determines the public's opinion on that issue; coverage that portrays affective imagery[8] or reflects certain value systems[9] may influence the policy opinions of viewers. As framing can differ from news source to news source, a consistent source of media can also drive policy opinions. Specifically, right wing sources often have steady messaging on policy areas, such as promoting climate change skepticism[10] or anti-immigrant sentiments.[11] Reaction to these framings, however, may vary based on existing policy beliefs and the policy area being presented.[12]

---

6 Kingdon, "Agendas, Alternatives, and Public Policies", 61
7 Gollust, Sarah E., Erika Franklin Fowler, and Jeff Niederdeppe. "Television News Coverage of Public Health Issues and Implications for Public Health Policy and Practice."
8 Leiserowitz, Anthony. "Climate Change Risk Perception and Policy Preferences: The Role of Affect, Imagery, and Values."
9 Solheim, Øyvind Bugge. "Are we all Charlie? How Media Priming and Framing Affect Immigration Policy Preferences After Terrorist Attacks."
10 Feldman, Lauren, Edward W. Maibach, Connie Roser-Renouf, and Anthony Leiserowitz. "Climate on Cable."
11 Hameleers, Michael. "Putting our Own People First: The Content and Effects of Online Right-Wing Populist Discourse Surrounding the European Refugee Crisis."
12 Ibid; Feldman, Lauren, Edward W. Maibach, Connie Roser-Renouf, and Anthony Leiserowitz. "Climate on Cable."

2.1.2… And Opinion As A Driver of Policy

As news media has been shown to drive public opinion and policy preferences, so too has public opinion been shown to be a driver of policy and legislative action. Linear regression has been a popular method for demonstrating the causal effect of opinion on policy. Research examining both energy policies in Europe[13] and abortion policy at the individual state level[14] shows the positive statistical significance of opinion, even when accounting for the existence of such policies. Furthermore, the influence of public opinion on policy at the state level is not limited to intrastate opinions. Political scientists have found that public opinion in neighboring states has shaped policy decisions of home states on issues of such as smoking[15] and the Affordable Care Act.[16]

Additional efforts to measure public opinion's effect on legislation were conducted through non-regression techniques. Through modeling "covariation" in survey policy data with changes in policy during the same time period that surveys were collected, researchers found that in a majority of instances of opinion change, policy changed with it.[17] Other research has shown graphically that varying types of antidiscrimination legislation passed after increases in public support and increases in national social movement action.[18]

---

13 Anderson, Brilé, Tobias Böhmelt, and Hugh Ward. "Public Opinion and Environmental Policy Output: A Cross-National Analysis of Energy Policies in Europe."
14 Kevin Arceneaux. "Direct Democracy and the Link between Public Opinion and State Abortion Policy."
15 Pacheco, Julianna. "The Social Contagion Model: Exploring the Role of Public Opinion on the Diffusion of Antismoking Legislation Across the American States."
16 Pacheco, Julianna and Elizabeth Maltby. "The Role of Public Opinion—Does it Influence the Diffusion of ACA Decisions?"
17 Page, Benjamin I., and Robert Y. Shapiro. "Effects of Public Opinion on Policy."
18 Burstein, Paul. "Public Opinion, Demonstrations, and the Passage of Antidiscrimination Legislation."

The relationship between media and public opinion is clear, especially when accounting for the source of media. As media drives public opinion, public opinion subsequently drives policymaking. Existing research, however, has largely focused on single issues or events; there is room to expand these studies to a more generalizable model. Given research specifying the importance of the source of news, H_1 can be updated to the following:

$H_{1a:}$ Arizona Capitol Times coverage of the topic of a bill prior to its introduction in the Arizona State Legislature increases the probability of that bill's probability of passage.

$H_{1b:}$ Arizona Daily Star coverage of the topic of a bill prior to its introduction in the Arizona State Legislature increases the probability of that bill's probability of passage.

$H_{1c:}$ Arizona Republic coverage of the topic of a bill prior to its introduction in the Arizona State Legislature increases the probability of that bill's probability of passage.

$H_{1d:}$ Phoenix New Times coverage of the topic of a bill prior to its introduction in the Arizona State Legislature increases the probability of that bill's probability of passage.

2.1.3 Additional Research into Policymaking

Rather than studying the connection from media to opinion and from opinion to policy, some political scientists have also sought to examine the more direct relationship between media and policymaking. In certain instances, research has found that media coverage of an issue has led to passage of legislation aimed at curtailing that issue.[19] Legislators also utilize the media to pass their own policy agenda; in their attempts at drawing support for

19 Douglas, Emily M. "Media Coverage of Agency-Related Child Maltreatment Fatalities: Does it Result in State Legislative Change Intended to Prevent Future Fatalities?"

the Patriot Act, the Bush Administration relied on appearances and writings in the media to sway support for their legislation.[20] The media provides a means by which the public can develop a base for their policy opinion development, and for involved actors such as legislators to influence such development.

Additional research has found the relationship between media, opinion, and policymaking to be less than straightforward. At the state-level, news coverage and public opinion have both been found to correlate with each other as well as with legislation[21]. In a similar analysis using simultaneous equations, political scientists have found that public opinion and legislation with regards to the Women's Movement were interdependent and predictive of one another.[22] Even in research seeking to demonstrate a more complicated relationship between the policymaking triad, the impact of media on policymaking still emerges as meaningful.

<u>2.2 Non Media Centric Methods of Understanding Policymaking</u>

H_2: Increases to bill meta-data variables, such as number of bill sponsors and number of committees introduced in, increases the probability of that bill passing the Legislature.

H_3: Different bill topics derived from bill text have statistically significant different effects on bill passage probability.

Researchers have generally focused on more traditional methods to explain and understand the legislative process. A popular theory has matched bill sponsorship to bill success. At the state-level, retirement bills with multiple sponsors and/or with sponsors

20 Domke, David, Erica S. Graham, Kevin Coe, Sue Lockett John, and Ted Coopman. "Going Public as Political Strategy: The Bush Administration, an Echoing Press, and Passage of the Patriot Act."
21 Tan, Yue and David H. Weaver. "Local Media, Public Opinion, and State Legislative Policies."
22 Costain, Anne N. and Steven Majstorovic. "Congress, Social Movements and Public Opinion: Multiple Origins of Women's Rights Legislation."

on committee chairs were found to pass a significantly higher rates than other bills.[23] Bill meta-data variables often capture political information about a bill; this information includes the total number of sponsors or the proportion of sponsors belonging to one political party. Additionally, federal bills with sponsors on the committee and/or matching the majority political party have been found to pass out of committees with greater success.[24]

Text analysis methods have also been found to be predictive of bill passage. Text-based models, coupled with bill meta-data such as bill sponsorship and sponsor ideology, have been found to improve the accuracy of models seeking to predict the passage of a bill.[25] These text models, however, have been built in the machine learning space without focus on a particular legislative area.

In seeking to predict the passage of bills in the Arizona State Legislature, any modeling efforts should include both bill meta-data variables and text data. Bill meta-data variables have shown to be inferential of bill movement at the state level, while text data has been shown to be predictive of all types of legislation.

2.3 Opportunity for Unification

Political scientists have long focused on quantifying the legislative process and its influences. Researchers have shown that media coverage of specific policy areas can lead to the passage of legislation and change in policy related to that specific area. However,

23 William P. Browne. "Multiple Sponsorship and Bill Success in U. S. State Legislatures."
24 Scott J. Thomas and Bernard Grofman. "Determinants of Legislative Success in House Committees."
25 S. M. Gerrish and D. M. Blei. "Predicting Legislative Roll Calls from Text."; Kornilova, Anastassia, Daniel Argyle, and Vlad Eidelman. "Party Matters: Enhancing Legislative Embeddings with Author Attributes for Vote Prediction"; John J. Nay. "Predicting and Understanding Law-Making with Word Vectors and an Ensemble Model."

that model has not been well studied as it relates to generalized legislation. Additionally, bill meta-data as well as bill text have been shown to enhance models aimed at predicting and understanding legislation. Given existing research, there is an opportunity to explore both generalized state-level media effects while unifying that exploration with separate research into the effects of bill text and of bill meta-data on bill success.

3. Data

The two primary sources of data for the model were Legiscan and Arizona based news sources. Legiscan contains many bill data sets for each bill introduced to each state's legislature. Some of these datasets include information for each step in the bill's history, for the sponsor of each bill, and for each bill's basic information (title, number, etc.). Arizona newspaper data – from The Arizona Capitol Times, The Arizona Republic, The Arizona Daily Star, and Phoenix New Times – was collected from each of their respective websites via python-based web-scraping algorithms developed by the author.

The key independent bill meta-variables derived from the Legiscan data are: the number of committees a bill was introduced to, the number of sponsors of a bill, and the proportion of Republican sponsors of a bill[26] for every bill introduced from 2015 through 2019. Additionally, Latent Dirichlet Allocation was used to reduce the size of the bill text data into three subtopics. The dependent variable of the study was a binary variable that described whether an introduced bill was chaptered into law.

[26] The Democratic Party has not held a branch of the Legislature in nearly 30 years (https://ballotpedia.org/Party_control_of_Arizona_state_government)

Researchers have typically made use of Lexis Nexis to find articles relating to a topic.[27] However, manual Lexis Nexis searching can be tedious if thousands of terms need to be searched. To get around this difficulty, an author-written web-scraping algorithms was used to search for articles using key terms and that limit results to articles written within a specified date range. Each search's key terms were the short title of each bill, and the date range was the six months leading to the input bill's introduction. Every introduced bill during the time frame of interest was run through each new site's respective scraper. The scraper ultimately returned the text of each article associated with each bill, and the count by news site that each bill's key word search returned.

4. Methods of Analysis

4.1 Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) is a probabilistic model for dimensionality reduction of text data.[28] With LDA, each unit of analysis (e.g. document or tweet) is assumed to contain multiple topics, and each word within that unit of analysis belongs to a topic.[29] An LDA outputs predicted probabilities that each topic is the primary topic within a document.[30]

27 Tan, Yue and David H. Weaver. "Local Media, Public Opinion, and State Legislative Policies."; Douglas, Emily M. "Media Coverage of Agency-Related Child Maltreatment Fatalities: Does it Result in State Legislative Change Intended to Prevent Future Fatalities?"; [27] Gollust, Sarah E., Erika Franklin Fowler, and Jeff Niederdeppe. "Television News Coverage of Public Health Issues and Implications for Public Health Policy and Practice."; Feldman, Lauren, Edward W. Maibach, Connie Roser-Renouf, and Anthony Leiserowitz. "Climate on Cable."
28 Grimmer and Stewart
29 Ibid
30 Ibid

4.2 Logistic Regression

A logistic regression was the model chosen to conduct the analysis. This generalized linear model, seen in Equation 1 below, estimates the probability that the binary dependent variable is equal to one.[31] In this equation, β_0 is the value of the intercept, β_n is the value of the coefficient for the *nth* variable in the model, and X_{ni} represents the value of the *nth* variable at its *ith* observation.[32]

$$Prob(Y_i = 1) = \frac{e^{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} \ldots + \beta_n X_{ni}}}{1 + e^{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} \ldots + \beta_n X_{ni}}}$$

Equation 1

While the coefficients associated with each variable did not have directly intuitive meaning, the coefficient's sign and magnitude depicts the degree to which each input affects the probability of that observation equaling one.

4.3 Log-likelihood Score

The measure of fit for both models will be the log-likelihood scores. Log-likelihood at its basic level is a measure of model fit. With respect to logistic regression, log-likelihood allows researchers to compare similar models to determine which combinations of variables best explain variation in the dependent variable. With respect to Latent Dirichlet Allocation, log-likelihood is a measure to compare models with different numbers of topics.[33] Essentially, log-likelihood measures how well the distribution of predicted topic probabilities match the true values. When comparing log likelihood scores of similar LDA models, the model with maximum scores is considered to have the best fit.

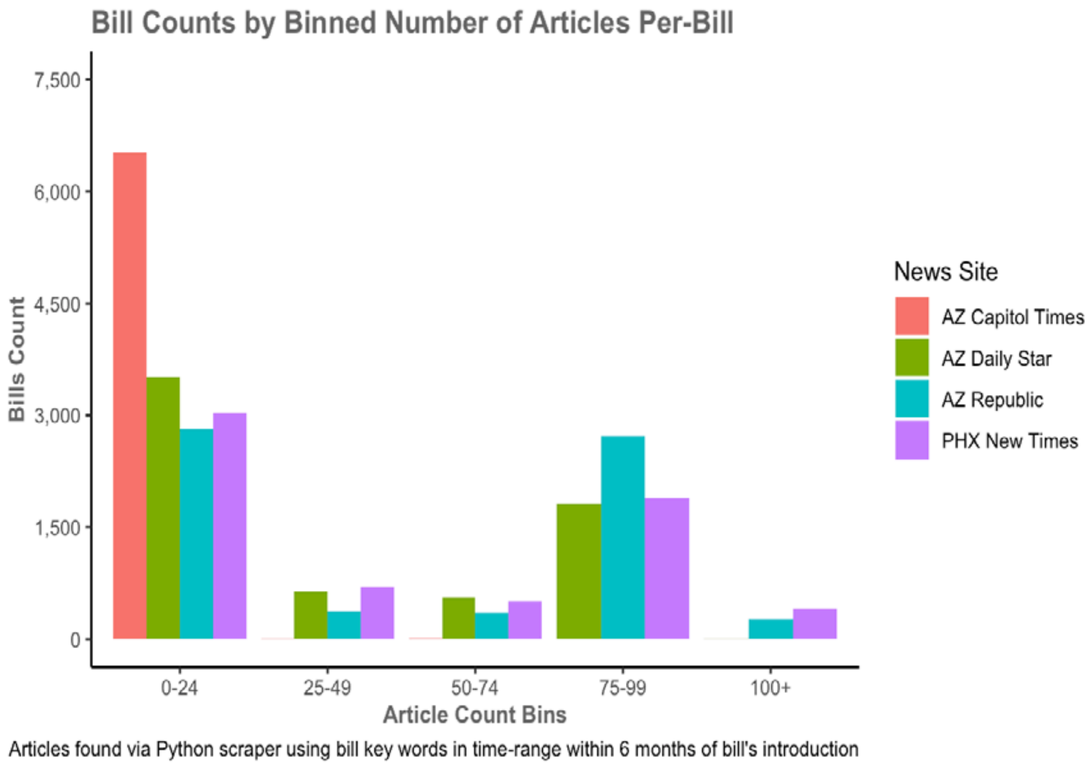
31 Bailey, Michael A. *Real Stats: Using Econometrics for Political Science and Public Policy*
32 Ibid
33 Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent Dirichlet Allocation."

5. Primary Variable Analysis

5.1 Article Counts

The article scrapers created the number of articles written about each bill's topic. Graph 1 groups the number of articles for a given bill into five bins, and then shows the number of bills that fit into each bin.

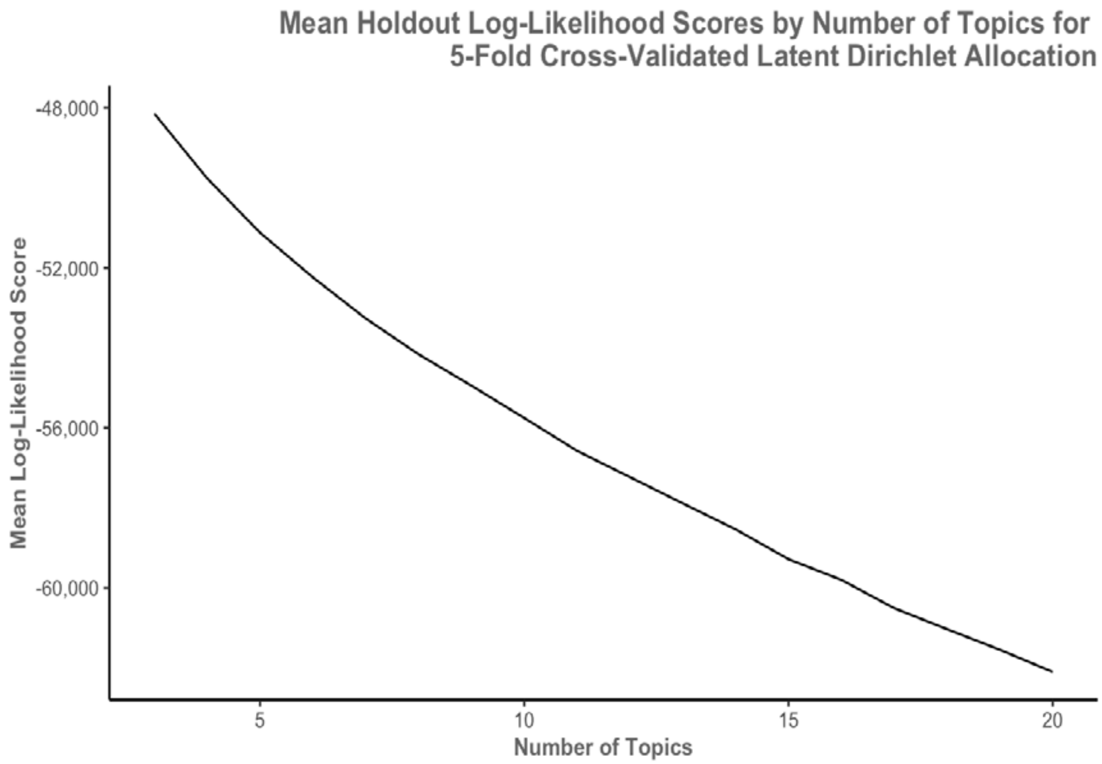
The above graph provides a couple of key insights. First, the Arizona Capitol Times does not produce many articles on a per-bill-topic basis. Second, the Arizona Republic and the Phoenix News are the only two news sites with a non-negligible number of associated bill topics with 100+ scraped articles. If the model supports the news hypotheses with high positive coefficients for The Arizona Republic and The Phoenix New Times, then this graph will provide supplementary evidence to support that bills with high article counts correlate to higher passage rates.



Graph 1

5.2 Bill Short Title Topic Model

The number of topics for the LDA was determined by comparing cross-validated log-likelihood scores of model results from a range of three to twenty topics. Graph 2 compares the average holdout log-likelihood score for each possible number of topics. With the goal finding the value closest to zero, the optimum number of topics based on the graph is three.



Graph 2

6. Model Results

6.1 Logistic Regression

Three logistic models were trained, and their results are shown in Table 1. Model 1 included article counts by news site and the bill topics as fit by the LDA. Model 2 added *Number of Committee Introductions* describing the number of committees a bill was introduced in and the binary variable *Senate Introduction* describing whether a bill was first introduced in the Senate (*Senate Introduction* = 1) or if a bill was first introduced in the House (*Senate Introduction* = 0). Model 3 added sponsorship variables *Number of Sponsors* describing the number of bill sponsors and *Percent Republican Sponsorship* describing the percentage of bill sponsors that were Republican. A factored year variable was included in every model to fix the temporal effects.

Logistic Regression Results

	Model 1	Model 2	Model 3
PHX New Times Article Count	0.0001 (0.001)	−0.0002 (0.001)	−0.0003 (0.001)
AZ Capitol Times Article Count	−0.007 (0.009)	−0.016 (0.010)	−0.007 (0.011)
AZ Republic Article Count	0.0001 (0.001)	0.0004 (0.001)	−0.001 (0.001)
AZ Daily Star Article Count	0.001 (0.001)	0.001 (0.001)	0.002* (0.001)
Number of Committee Introductions		1.325*** (0.050)	1.101*** (0.055)
Senate Introduction		0.015 (0.061)	0.208*** (0.066)
Number of Sponsors			0.026*** (0.005)
Percent Republican Sponsorship			3.520*** (0.155)
Topic2	−0.354*** (0.072)	−0.218*** (0.077)	−0.278*** (0.084)
Topic3	−0.019 (0.066)	0.109 (0.071)	0.045 (0.078)
2016	0.116 (0.098)	0.083 (0.106)	0.258** (0.112)
2017	0.123 (0.093)	0.239** (0.099)	0.299*** (0.106)
2018	−0.008 (0.092)	0.082 (0.099)	0.294*** (0.106)
2019	−0.192** (0.093)	0.034 (0.100)	0.363*** (0.108)
Constant	−0.987*** (0.088)	−3.296*** (0.134)	−6.028*** (0.211)
N	6540	6540	6540
Log Likelihood	−3728.178	−3327.545	−2771.271
AIC	7478.355	6681.089	5572.543

***p < .01; **p < .05; *p < .1

Table 1

Looking at the models, Model 3 appears to fit the data best given that it has the highest log-likelihood score of the three models. The model coefficients for the article count variables all fail to achieve statistical significance, with the exception of Model 3 where the number of articles written about a bill topic from the Arizona Daily Star show a positive, statistically significant impact on bill passage rate. As the coefficient of the Arizona Daily Star is positive, then Model 3 supports H_{1b} while rejecting H_{1a}, H_{1c}, and H_{1d}. Additionally, bills with short titles sorted into Topic 2 have statistically significantly lower probabilities of passing than those sorted into Topic 1; bills with short titles sorted into Topic 3 pass at the statistically similar rates as those sorted into Topic 1. The bill meta-variables, Committee Introduction Count, Senate Introduction, Sponsor Count, and Republican Sponsorship Percentage, all displayed statistical significance associated with positive influence on bill passage.

6.2 Difference in Estimated Effects

To better understand the substantive significance of the statistically significant variable coefficients, differences in fit probabilities were calculated. The purpose of these calculations is to understand substantive impact of variables deemed statistically significant; if there is no meaningful difference between two extremes of a variable, then the statistical significance is unimportant compared to its substantive significance.

The maximums and minimums of numerical statistically variables were fit to Model 3 and then compared for substantive significance. For the bill topic variable, Topic1 and Topic2 were fit to Model 3 and compared for substantive significance. The Senate Introduction binary variable was fit to Model 3 for values of 1 and 0, and compared for substantive significance. When performing these comparisons, all numerical variables of non-interest were set to their mean, the Topic variable was set to Topic1 and Senate Introduction was set to 0.

Variable	Passage Probability Difference	Passage Probability 1	Passage Probability 2
Committee Introduction Count	80.02%	10.85%	90.87%
Sponsor Count	43.21%	15.97%	59.17$
Republican Sponsorship Percentage	42.2%	2.32%	44.52%
Topic	-3.66%	17.55%	13.89%
AZ Daily Star Count	2.85%	16.57%	19.42%
Senate Introduction	2.81%	14.74%	17.55%

Table 2

In Table 2 above, the *Variable* column represents the variable of interest; the *Passage Probability Difference* represents the differences in fit probabilities between *Probability 1* and *Probability 2*; *Probability 1* represents the predicted probability when a numeric variable of interest is set to its minimum, the Topic variable is set to Topic1, or the Senate Introduction variable is set to 0; and *Probability 2* represents when a numeric variable of interest is set to its maximum, the Topic variable is set to Topic 2, or the Senate Introduction variable is set 1.

7. Discussion

7.1 Modeling Results

The initial model results support all three hypotheses. The Arizona Daily Star coverage of a bill topic positively influences the probability of a bill passing the Arizona State Legislature. Bill meta data, captured by the two introduction and two sponsorship variables, was also shown to be statistically significant and positively correlate with the passage rate of a bill. The text of the bill title also contained information that is inferential of bill passage probability.

Despite these instances of statistical significance in the variables of interest, the observed effects of changes to those variables provided more insight into their effects on bill passage probability. Comparing the maximum and minimums of bill meta variables Committee Introduction Count, Sponsor Count, and Republican Sponsorship percentage show substantive changes in passage probability. The comparisons of changes to Senate Introduction, bill short title Topic, and the Arizona Daily Star article count did not show substantive differences in the probability of a bill passing. Overall, the observed differences allow for the rejection of hypotheses one and three, given that the variables associated with these hypotheses displayed absolute probability passages difference of less than 5% in the observed value exercise.

7.2 Theoretical Significance

The rejection of the H_{1X} hypotheses connects to existing theory in important ways.

Previous research has shown that media drives singular policy topics on a national scale and that coverage on a policy area can lead to the passing of legislation related to that topic. This research, however, has failed to demonstrate a similar media effect on a local scale and without focus on specific policy topic. The acceptance of H_2 allows for the expansion of existing research. Previous studies have shown that bill meta-variables positively impact the movement of bills through the legislative process at federal and policy-specific levels; this study expands that research to state-level, generalizable bill topics.

The results relating to the text hypothesis, when compared to the other hypotheses, are not as straightforward. Most existing research has focused on the predictive power of bill text more so than its effects on bill passage. This research shows that grouping bill topics together does not provide significant insight into bill passage rate at a generalizable, state level scale. There is room, however, to follow the example set by media effects research and focus text analysis on single topic at the national scale or include a wider range of topics.

8. Conclusion

This research primarily sought to prove a positive relationship between the coverage of a bill's topic in Arizona-based news organizations and the passage of that bill in the Arizona State Legislature. The results of the research, however, failed to show that generalized, state-level coverage of a bill impacts that bill's passage probability. In the context of existing research, these results shed light on their key finding. Previous studies have shown that media can impact specific policy areas on national, international, and multi-state levels; the observed impact cannot be extended to a generalizable, single-state level.

The additional hypotheses also contribute to previous studies. Text analysis methods have largely been used to increase the predictive power of machine learning models; this research fails to prove that text data has inferential power in understanding the probability of bill passage. Grouping bill summary text into a small number of groups may simply not capture the complexity of text data.

Bill meta data has been shown, when focusing on specific committees and policy areas, to positively impact bill passage. This research shows more bill sponsors, higher proportions of majority party sponsors, and more committee introductions all increase the probability of a bill passing the legislature. These results expand existing research by demonstrating the effect of bill meta data on generalized legislation.

Future research efforts should primarily seek to build on two aspects of this research: the incorporation of media effects and the scope of the topic modeling. The article search algorithm does a good job of scraping articles by key word within a certain timeframe. What the algorithms ultimately lack that a service like Lexis Nexis does not, however, is an ability to apply relevance scoring to the articles. Either filtering on relevance score or taking a weighted mean of article counts by relevance score would most likely improve the validity of the model results. In addition to relevance score, incorporating article sentiment analysis could also add more contextual information for how a certain bill is covered. Finally, it would be worthwhile to determine if there is multicollinearity between the article search variables that would otherwise weaken the statistical impact of such variables.

The scope of the text topic modeling can also be improved upon for future research efforts. As has been noted, past studies have incorporated a much larger volume of text into predictive modeling efforts. Due to time and resource limitations, this study could only incorporate the few words per bill short title as opposed to the hundreds of words contained in complete or summary bill text. Modeling full bill text and the scraping of bill summaries were beyond the scope of this study but should be examined in future research efforts.

Bibliography

Anderson, Brilé, Tobias Böhmelt, and Hugh Ward. "Public Opinion and Environmental Policy Output: A Cross-National Analysis of Energy Policies in Europe." *Environmental Research Letters* 12, no. 11 (2017): 114011, https://search.datacite.org/works/10.1088/1748-9326/aa8f80.

Bailey, Michael A. *Real Stats: Using Econometrics for Political Science and Public Policy*. 1st ed. New York, NY: Oxford University Press (2016).

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent Dirichlet Allocation." Journal of Machine Learning Research, (2004). https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf.

Burstein, Paul. "Public Opinion, Demonstrations, and the Passage of Antidiscrimination Legislation." *Public Opinion Quarterly* 43, no. 2 (1979): 157-172, https://search.datacite.org/works/10.1086/268508.

Costain, Anne N. and Steven Majstorovic. "Congress, Social Movements and Public Opinion: Multiple Origins of Women's Rights Legislation." *Political Research Quarterly* 47, no. 1 (Mar, 1994): 111.

Domke, David, Erica S. Graham, Kevin Coe, Sue Lockett John, and Ted Coopman. "Going Public as Political Strategy: The Bush Administration, an Echoing Press, and Passage of the Patriot Act." *Political Communication* 23, no. 3 (Sep 1, 2006): 291-312, http://www.tandfonline.com/doi/abs/10.1080/10584600600808844.

Douglas, Emily M. "Media Coverage of Agency-Related Child Maltreatment Fatalities: Does it Result in State Legislative Change Intended to Prevent Future Fatalities?" *Journal of Policy Practice* 8, no. 3 (Jun 11, 2009): 224-239, http://www.tandfonline.com/doi/abs/10.1080/15588740902963445.

Feldman, Lauren, Edward W. Maibach, Connie Roser-Renouf, and Anthony Leiserowitz. "Climate on Cable." *The International Journal of Press/Politics* 17, no. 1 (Jan, 2012): 3-31, https://journals.sagepub.com/doi/full/10.1177/1940161211425410.

Grimmer, Justin and Brandon M. Stewart. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." Political Analysis 21, no. 3 (2013): 267-297, https://dx.doi.org/10.1093/pan/mps028.

Gollust, Sarah E., Erika Franklin Fowler, and Jeff Niederdeppe. "Television News Coverage of Public Health Issues and Implications for Public Health Policy and Practice." *Annual Review of Public Health* 40, no. 1 (Apr 1, 2019): 167-185, https://www.ncbi.nlm.nih.gov/pubmed/30633711.

Hameleers, Michael. "Putting our Own People First: The Content and Effects of Online Right-Wing Populist Discourse Surrounding the European Refugee Crisis." *Mass Communication & Society* 22, no. 6 (2019): 804-826, https://search.datacite.org/works/10.1080/15205436.2019.1655768.

John J. Nay. “Predicting and Understanding Law-Making with Word Vectors and an Ensemble Model.” *PLoS ONE 12(5): e0176999* (July 6, 2016)

Justice, G. (2015, January 25). “States Six Times More Productive Than Congress”, Connectivity. https://info.cq.com/resources/states-six-times-more-productive-than-congress/

Kevin Arceneaux. "Direct Democracy and the Link between Public Opinion and State Abortion Policy." *State Politics & Policy Quarterly* 2, no. 4 (Dec 1, 2002): 372-387, https://www.jstor.org/stable/40421471.

Kingdon, John W. *Agendas, Alternatives, and Public Policies*. New York: Longman, (1995).

Kornilova, Anastassia, Daniel Argyle, and Vlad Eidelman. *Party Matters: Enhancing Legislative Embeddings with Author Attributes for Vote Prediction* (2018), https://www.aclweb.org/anthology/P18-2081.pdf

Leiserowitz, Anthony. "Climate Change Risk Perception and Policy Preferences: The Role of Affect, Imagery, and Values." *Climatic Change* 77, no. 1-2 (2006): 45-72, https://search.datacite.org/works/10.1007/s10584-006-9059-9.

McCombs, Maxwell E. and Donald L. Shaw. "The Agenda-Setting Function of Mass Media." *Public Opinion Quarterly* 36, no. 2 (1972): 176-187, https://search.datacite.org/works/10.1086/267990.

Pacheco, Julianna. "The Social Contagion Model: Exploring the Role of Public Opinion on the Diffusion of Antismoking Legislation Across the American States." *The Journal of Politics* 74, no. 1 (Jan, 2012): 187-202, https://dx.doi.org/10.1017/S0022381611001241.

Pacheco, Julianna and Elizabeth Maltby. "The Role of Public Opinion—Does it Influence the Diffusion of ACA Decisions?" *Journal of Health Politics, Policy and Law* 42, no. 2 (Apr, 2017): 309-340, http://dx.doi.org/10.1215/03616878-3766737.

Page, Benjamin I., and Robert Y. Shapiro. "Effects of Public Opinion on Policy." American Political Science Review 77, no. 1 (1983): 175-90. doi:10.2307/1956018.

Parrott, Scott, Jennifer Hoewe, Minghui Fan, and Keith Huffman. "Portrayals of Immigrants and Refugees in U.S. News Media: Visual Framing and its Effect on

Emotions and Attitudes." *Journal of Broadcasting & Electronic Media* 63, no. 4 (2019): 677-697, https://search.datacite.org/works/10.1080/08838151.2019.1681860.

Pew Research Center. “Cable News Fact Sheet.” Washington, D.C. (June 25, 2019) https://www.journalism.org/fact-sheet/cable-news/

Pew Research Center. “Local TV News Fact Sheet.” Washington, D.C. (June 25, 2019) https://www.journalism.org/fact-sheet/local-tv-news/

Pew Research Center. “Network News Fact Sheet.” Washington, D.C. (June 25, 2019) https://www.journalism.org/fact-sheet/network-news/

Rosen, J. “Americans Don't Know Much about State Government”*, Survey Finds*. (2018, December 14) https://hub.jhu.edu/2018/12/14/americans-dont-understand-state-government/

Scott J. Thomas and Bernard Grofman. "Determinants of Legislative Success in House Committees." *Public Choice* 74, no. 2 (Jan 1, 1992): 233-243, https://www.jstor.org/stable/30025600.

Solheim, Øyvind Bugge. "Are we all Charlie? how Media Priming and Framing Affect Immigration Policy Preferences After Terrorist Attacks." *West European Politics* (2019): 1-25, https://search.datacite.org/works/10.1080/01402382.2019.1683791.

S. M. Gerrish and D. M. Blei. “Predicting Legislative Roll Calls from Text.” *In Proceedings of the 28th Annual International Conference on Machine Learning, ICML '11* (June 2011) 489-496 https://dl.acm.org/doi/10.5555/3104482.3104544

Tan, Yue and David H. Weaver. "Local Media, Public Opinion, and State Legislative Policies." *The International Journal of Press/Politics* 14, no. 4 (Oct, 2009): 454-476, https://journals.sagepub.com/doi/full/10.1177/1940161209336225.

William P. Browne. "Multiple Sponsorship and Bill Success in U. S. State Legislatures." *Legislative Studies Quarterly* 10, no. 4 (Nov 1, 1985): 483-488, https://www.jstor.org/stable/440070.

Appendix

Item 1: Legiscan information

Legiscan data can be accessed by creating a free account. Next, state level .csv files can be downloaded by year. The data is contained is a single folder entitled "csv", and contains the following files: bills.csv, documents.csv, history.csv, people.csv, rollcalls.csv, sponsors.csv, votes.csv. A README.md file is also contained in the csv folder, and contains bill column names and descriptors for each variable in each dataset.

Item 2: Key words by bill topic

The below table contains key words in each of the three modeled bill topics.

Model Topics	Key Word 1	Key Word 2	Key Word 3	Key Word 4	Key Word 5	Key Word 6	Key Word 7
Topic 1	appropriation	service	registration	right	arizona	veteran	fund
Topic 2	correction	technical	repeal	vehicle	insurance	medial	tax
Topic 3	school	tax	public	education	program	board	budget

Table 3

Item 3: Bill Text Scraping Chunk (Python)

Below is a chunk of code used to scrape the bill text. The complete set of code used for the entire project can be found at:

```
#create list for each variable
years = []
bill_numbers = []
bill_texts = []

#time how long it takes for entire loop and for each session
total_time = time.time()

#loop through each year
for session in directories:
   session_start = time.time()
   #get year of bill
   year = session[-4:]
   #read in necessary files
   documents = pd.read_csv(session + '/csv/documents.csv')
   bills = pd.read_csv(session + '/csv/bills.csv')

   #merge dfs together to get bill number
   bills_select = bills[['bill_id', 'bill_number', 'last_action_date']]
   merged = pd.merge(bills_select, documents, on =
'bill_id').sort_values('last_action_date').drop_duplicates('bill_id',keep='last').reset_ind
ex()

   #create directory to save bill text by year if none exists
   directory = '../Data/BillText/' + year + '/'
   if not os.path.exists(directory):
      os.makedirs(directory)

   #iterate through each bill, save a txt file
   for i in range(len(merged)):
      time.sleep(5)
      print(i)
      if os.path.isfile(directory + merged['bill_number'][i] + '.txt'):
         continue
      else:
         #save url
         url = merged['url'][i]

         #request url
         page = requests.get(url, verify=False)
```

```
        #parse html
        soup = BeautifulSoup(page.content, 'html.parser')

        #get bill text
        text = soup.find_all('div', attrs = {'class': 'WordSection2'})
        if len(text) != 0:
          for whole_bill in text:
              main_bill = whole_bill.find_all('p')
              bill_text = [paragraph.text for paragraph in main_bill if paragraph.text !=
'']
        else:
          text = soup.find_all('div', attrs = {'class': 'Section2'})
          for whole_bill in text:
            main_bill = whole_bill.find_all('p')
            bill_text = [paragraph.text for paragraph in main_bill if paragraph.text !=
'']
        #create single string
        bill_text = ' '.join(bill_text)

        #write txt file
        file = open(directory + merged['bill_number'][i] + '.txt', "w")
        file.write(bill_text)
        file.close()

        #append values to lists to be added to dataframe
        years.append(year)
        bill_numbers.append(merged['bill_number'][i])
        bill_texts.append(bill_text)
```

Item 4: Article Scraping Chunk (Python)

Below is a chunk of code from one of the Arizona Daily Star article scraping functions. The full code for each article scraping script can be found in the Github Repository mentioned above.

```
page = requests.get(url)

    #parse html
    soup = BeautifulSoup(page.content, 'html.parser')

    #get title text
    title_tag = soup.find('h1', attrs = {'class':'headline'})
    title_text = title_tag.find_all('span')
    titles = [title.text.strip() for title in title_text]
```

```
    #get author name
    author_list = []
    author_tag = soup.find('span', attrs = {'itemprop':'author'})
    author_list.append(author_tag.text.strip())

    author_list = [author.replace('By ', '') for author in author_list]
    author_list = [author.split("\n",2)[0] for author in author_list]



    #get article text
    text = soup.find_all('p', attrs = None)

    articles = [paragraph.text for paragraph in text if paragraph.text != '']

    articles = ' '.join(articles)

    drop = "To continue viewing content on tucson.com, please sign in with your
existing account or subscribe. We have not been able to find your subscription. Current
Subscriber? Log in Current Subscriber? Activate now Or Don't have a subscription?
Subscribe now Subscribe today for unlimited access Subscribe today for unlimited
access"

    articles = articles.replace('\n', '').replace('\t', '').replace(drop, '')

    return(titles[0], author_list[0], articles)
```

Item 5: Topic Modeling (Python):

Below is the code that was used to develop the bill title topics. The full code for each

article scraping script can be found in the Github Repository mentioned above.

```
#initialize empty dataframe

df = pd.DataFrame(columns = ['bill_number', 'year', 'title'])
# create complete df of bill titles
for year in years:
  #read in year bills df
  bill_df = pd.read_csv('../Data/Legiscan/' + str(year) + '/csv/bills.csv')

  #create year column as another key
  bill_df['year'] = str(year)
```

```
    #select columns of interest
    bill_df_select = bill_df[['bill_number', 'year', 'title']]

    #concat to df of interest
    df = pd.concat([df, bill_df_select], axis = 0)

# Initialize the Wordnet Lemmatizer
def lemmatize_text(text):
    '''
    Function that lemmatizes text

    Input:
    -String

    Output:
    -Lemmatized string
    '''
    #initialize lemmatizer
    words = nlp(' '.join(text))

    #return list of lemmatized text
    return([word.lemma_ for word in words])

df = df.reset_index(drop = True)

#convert single string to list
df['title'] = df['title'].apply(lambda x: x.lower().split('; '))
df.title = df.title.apply(lambda x: ' '.join(x).split(' '))

#remove stopwords
df['title'] = df['title'].apply(lambda x: [item for item in x if item not in
stopwords.words('english')])

#lemmatize text
df['title'] = df['title'].apply(lambda x :lemmatize_text(x))

df['title'] = df['title'].apply(lambda x: [word for word in x if not word[0].isdigit()])
df['title'] = df['title'].apply(lambda x: [word for word in x if word not in
string.punctuation])

#join text
df.title = df.title.apply(lambda x: ' '.join(x))

#initialize vectorizer
vectorizer = CountVectorizer()
```

```
#vectorize column
title_vectors = vectorizer.fit_transform(df['title'])


### Topic Modeling ###

# grid search set up

# optimal number of parameters, min number of committees to max number of
committees
topics = np.arange(3, 21, 1)

# learning rate
learning_rate = np.arange(0.5, 0.9, 0.1)

#buil param dict
params = {'n_components':topics, 'learning_decay':learning_rate}


# initialize
lda_gridsearch = LDA(random_state = 1234)

# Init Grid Search Class
model_search = GridSearchCV(lda_gridsearch, param_grid = params, cv = 5)

#fit model
model_search.fit(title_vectors)

#pull out best paramters
best_lda_params = model_search.best_params_

#train best model
best_lda = LDA(random_state = 1234, n_components =
best_lda_params['n_components'],
        learning_decay = best_lda_params['learning_decay'])

#fit best model
best_lda.fit(title_vectors)

#transform titles into topics
topic_fit = best_lda.transform(title_vectors)

# column names
topicnames = ["Topic" + str(i + 1) for i in range(best_lda.n_components)]

# Make the pandas dataframe
```

```
df_topics = pd.DataFrame(np.round(topic_fit, 2), columns=topicnames)

# join with bill title data
df_topics = pd.concat([df.reset_index(drop=True), df_topics.reset_index(drop=True)],
axis = 1)

#save as csv
df_topics.to_csv('../Data/BillTitleTopics.csv')
```

Item 6: Data Merging (R)

Below is a chunk from the code that merges the bill demographic variables, article text data variables, and bill text data variables together. The full code can be found in the Github repository mentioned above.

```
## Arizona Daily Star ##
azds = NULL

for (year in years) {
  file = paste0('../Data/ArticleText/azds_articletext_', as.character(year), '.csv')
  if (is.null(azds)) {
    azds = read.csv(file)
  } else {
    azdsyear = read.csv(file)
    azds = rbind(azds, azdsyear)
    remove(azdsyear)
  }
}

azds$X = NULL

#remove duplicates
azds = azds %>% distinct(Author, Bill, Text, Title, Year)

#summarise data to create average by bill, year
azds_sum = azds %>%
  group_by(Bill, Year) %>%
  summarise(AZDS_Count = n())

### bills ###
#read in bills to get info if bill passed or not
bill_status = data.frame()
for (year in 2015:2019) {
  #convert to year to read in bill
```

```
 year = as.character(year)
 #create path variable
 path= paste0("../Data/Legiscan/", year, "/csv/")
 #read in bills csv and prep it
 year_df = read.csv(paste0(path, 'bills.csv'))
 year_df$passed = ifelse(str_detect(year_df$last_action, 'Chapter') == T, 1, 0)
 year_select = year_df %>% select(bill_number, passed, bill_id)
 year_select$year = year

 #read in hist csv
 hist = read.csv(paste0(path, 'history.csv'))
 hist$action = as.character(hist$action)

 hist_select = hist %>%
   group_by(bill_id) %>%
   #only keep actions where bills is assigned to committe
   filter(str_detect(action, pattern = 'Assigned')) %>%

   #remove rules committe(since every bill is assigned to Rules), unless Rules is only
committee of assignment
   mutate(count = n(),
        sen_intro = ifelse(str_detect(action, 'Senate') == T, 1, 0)) %>%

   filter((str_detect(action, pattern = 'RULES', negate = T)) | (str_detect(action, pattern
= 'RULES') & count == 1)) %>%

   #remove other useless words & create variable if introduced in senate
   mutate(action = removeWords(action, c('Assigned to ', ' Committee'))) %>%

   #filter so that only house of initial introduction is kept
   filter((sequence - min() < 5) & ((sen_intro == 1 & str_detect(action, 'Senate')) |
(sen_intro == 0 & str_detect(action, 'House')))) %>%

   #remove Senate and House from actions
   mutate(action = str_remove(action, 'Senate'),
        action = str_remove(action, 'House')) %>%

   #create concatenated variable for committees introduced in
   mutate(Committees_Concat = paste(unique(action), collapse = ','),
        Committee_Intro_Count = n()) %>%

   #only keep unique rows
   distinct(bill_id, Committees_Concat, Committee_Intro_Count, sen_intro) %>%

   #only choose first
   filter(row_number() == 1) %>%
```

```
  #keep only what we need
  select(bill_id, Committee_Intro_Count, Committees_Concat, sen_intro)

 #get sponsor of bill
 people = read.csv(paste0(path, 'people.csv'))
 people_select = people %>%
          #binary variable for party of sponsor
          mutate(RepublicanSponsor = ifelse(party == 'R',1,0)) %>%
          #keep variables of interest
          select(people_id, RepublicanSponsor)

 #add sponsor data
 sponsors = read.csv(paste0(path, 'sponsors.csv'))

 #merge together
 merg = merge(sponsors, people_select)
 #count bill sponsors
 grp = merg %>%
     group_by(bill_id) %>%
     summarise(SponsorCount = n(),
          RepubSponsorPct = round(sum(RepublicanSponsor)/n(), 3))

 hist_select = merge(hist_select, grp, on = 'bill_id', all.x = T)

 #merge two dfs together
 year_select = merge(year_select, hist_select, on = 'bill_id', all.x = T)

 bill_status = rbind(bill_status, year_select)
}

#fill in nas
bill_status = bill_status %>%
 mutate(Committee_Intro_Count = na.fill(Committee_Intro_Count, 0),
     Committees_Concat = na.fill(Committees_Concat, 'RULES'),
     sen_intro = ifelse(str_sub(bill_status$bill_number, 1, 1) == 'S', 1, 0))

#merge in article data
model_df = merge(model_df, azds_sum,
        by.x = c('bill_number', 'year'), by.y = c('Bill', 'Year'), all.x = T) %>%
     mutate(AZDS_Count = na.fill(AZDS_Count, 0))
```

Curriculum Vita

The author was born in 1995 and raised in New Jersey. He received his Bachelor's of Science in Political Science at Arizona State University in May of 2018, graduating with *magna cum laude* honors. His undergraduate thesis, entitled "Comparing the Influence of True/False Information on Climate Policy Preferences," was a survey experiment that sought to understand the effect of information on policy preferences. In August of 2018, the author began his Masters' education at Georgetown University, entering the Data Science for Public Policy program. After two semesters, he transferred to John Hopkins' M.S. Data Analytics for Policy program for non-academic reasons. He is graduating in December 2020 upon the submission of his capstone.